PUBLISHED IN SPIE PROCEEDINGS, WAVELET APPLICATIONS, ORLANDO, FL, APRIL, 1994.

Denoising and Robust Non-Linear Wavelet Analysis

Andrew G. Bruce
StatSci Division, MathSoft
1700 Westlake Ave. N,
Seattle, WA 98109

David L. Donoho
Department of Statistics
Stanford University
Stanford, CA 94305

Hong-Ye Gao
StatSci Division, MathSoft
1700 Westlake Ave. N,
Seattle, WA 98109

R. Douglas Martin
StatSci Division, MathSoft
1700 Westlake Ave. N,
Seattle, WA 98109

ABSTRACT

In a series of papers, Donoho and Johnstone develop a powerful theory based on wavelets for extracting non-smooth signals from noisy data. Several nonlinear smoothing algorithms are presented which provide high performance for removing Gaussian noise from a wide range of spatially inhomogeneous signals. However, like other methods based on the linear wavelet transform, these algorithms are very sensitive to certain types of non-Gaussian noise, such as outliers. In this paper, we develop *outlier resistant* wavelet transforms. In these transforms, outliers and outlier patches are localized to just a few scales. By using the outlier resistant wavelet transforms, we improve upon the Donoho and Johnstone nonlinear signal extraction methods. The outlier resistant wavelet algorithms are included with the S+WAVELETS object-oriented toolkit for wavelet analysis.

1 INTRODUCTION

The introduction of wavelets in the late 1980's has spawned a flurry of research activity, exploring new techniques for analysis of data simultaneously in the time and frequency domains. Several new "wavelet-like" transforms have been developed, such as wavelet packets, local cosine bases, Wilson bases, and matching pursuits. Wavelets have proven valuable for a variety of statistical applications, such as the denoising of signals or estimation of spectral or probability densities.

The presence of outliers in data causes problems in traditional time series analysis techniques. Outliers can seriously distort the autocorrelation function, partial autocorrelation function, spectral density function, model identification, and parameter estimates for models. Outliers can also cause problems with methods based on the wavelet decomposition. Wavelets are a linear transformation of the data, and hence, outliers have unbounded influence on the wavelet coefficients.

In this paper, we review research into new robust wavelet decompositions which are designed for analysis of data which contains outliers. Based on these decompositions, we extend wavelet-based statistical algorithms to handle a broader class of problems. In particular, we focus on the robust "smoother-cleaner" wavelet decomposition. Smoother-cleaner wavelets are an adaptation of the pyramid algorithm in which outliers captured into robust residuals at different multiresolution levels. The algorithm is computationally very fast with $O(n)$ complexity.

The paper is organized as follows. Section 2 reviews the wavelet-based denoising procedure of Donoho and Johnstone. Section 3 motivates the need for new robust wavelet decompositions, and presents decompositions based on minimizing norms which are insensitive to outliers. These decompositions have nice theoretical properties but are computationally slow. Section 4 presents the robust smoother-cleaner wavelet algorithm. The algorithm is applied to simulated data and radar glint noise data. Finally, section 5 gives a discussion of related research. This includes research into other robust wavelet decompositions, and the development S+WAVELETS, an object-oriented toolkit for wavelet analysis.

2 DENOISING BY WAVELET SHRINKAGE

Suppose our data x_i are noisy samples from a function f:

$$x_i = f(i/n) + \epsilon_i \tag{1}$$

where ϵ_i are iid $N(0, \sigma^2)$. We want to find an estimate $\hat{f}$ which minimizes the risk $R(\hat{f}, f) = E \parallel \hat{f} - f \parallel_2^2$. In a series of papers [6, 8, 7], Donoho and Johnstone propose a collection of related techniques which solve this problem. Their denoising procedure, which we refer to as WAVESHRINK, is based on a theoretically motivated nonlinear shrinkage of wavelet coefficients. The principle is that noise contributes to many coefficients but features contribute to only a few coefficients. Hence, by setting the smaller coefficients to zero in a statistically guided manner, we can nearly optimally eliminate noise while preserving the underlying signal.

The three steps in the WAVESHRINK algorithm are

[1] Apply the wavelet transform with J levels to the signal $\mathbf{X}$, obtaining wavelet detail and smooth coefficients D(1), D(2), ..., D(J), S(J).

[2] Shrink the detail coefficients at the j finest scales to obtain new detail coefficients $\tilde{\mathrm{D}}(1) = \delta_{\lambda_1}(\mathrm{D}(1)), \ldots, \tilde{\mathrm{D}}(j) = \delta_{\lambda_j}(\mathrm{D}(j))$. A statistically attractive form for the thresholding function is a *soft threshold*:

$$\delta_{\lambda_i}(x) = \begin{cases} 0 & \text{if } |x| \le \lambda_i \\ \mathrm{sign}(x)(|x| - \lambda_i) & \text{if } |x| > \lambda_i \end{cases} \tag{2}$$

Note that the threshold λ_i may vary from level to level.

[3] Apply the inverse wavelet transform to obtain the estimated smooth $\hat{\mathbf{X}}$.

Theoretical results show that for certain choices of the λ_j, the WAVESHRINK estimate $\hat{f}_{ws}$ can achieve nearly the *minimax* risk over a broad class of functions $\mathcal{F}$:

$$R(\hat{f}_{ws}, f) \approx \inf_{\hat{f}} \sup_{f \in \mathcal{F}} R(\hat{f}, f) \tag{3}$$

A consequence of this result is that the WAVESHRINK algorithm has a *locally adaptive bandwidth*. It has been shown to perform remarkably well on a broad range of spatially inhomogeneous signals. The smooth is completely automatic: no tuning constants are required (other than the choice of the wavelet filter and thresholding rule). The WAVESHRINK algorithm can be extended to other orthonormal bases as well, such as wavelet packets and local cosine bases[5].

3 ROBUST WAVELET DECOMPOSITIONS

The theory of Donoho and Johnstone demonstrates that wavelets provide a powerful framework for denoising data. However, this theory is based on the assumption that the noise ϵ_i is close to a Gaussian distribution. As a result, the WAVESHRINK algorithm is very sensitive to outliers.

Figure 1 compares WAVESHRINK for an artificial signal contaminated with Gaussian and non-Gaussian noise. Figure 1(a) displays the "`jumpsine`" signal: a sinusoid with a jump in the middle. The adjacent plot is the wavelet decomposition of the `jumpsine` signal. All of the large coefficients at the finer levels correspond to the jump. Figure 1(b) gives the signal plus Gaussian noise with the WAVESHRINK smooth. In this example, WAVESHRINK performs well and the smooth is very close to the original signal. This smooth is derived by inverting the "shrunken"

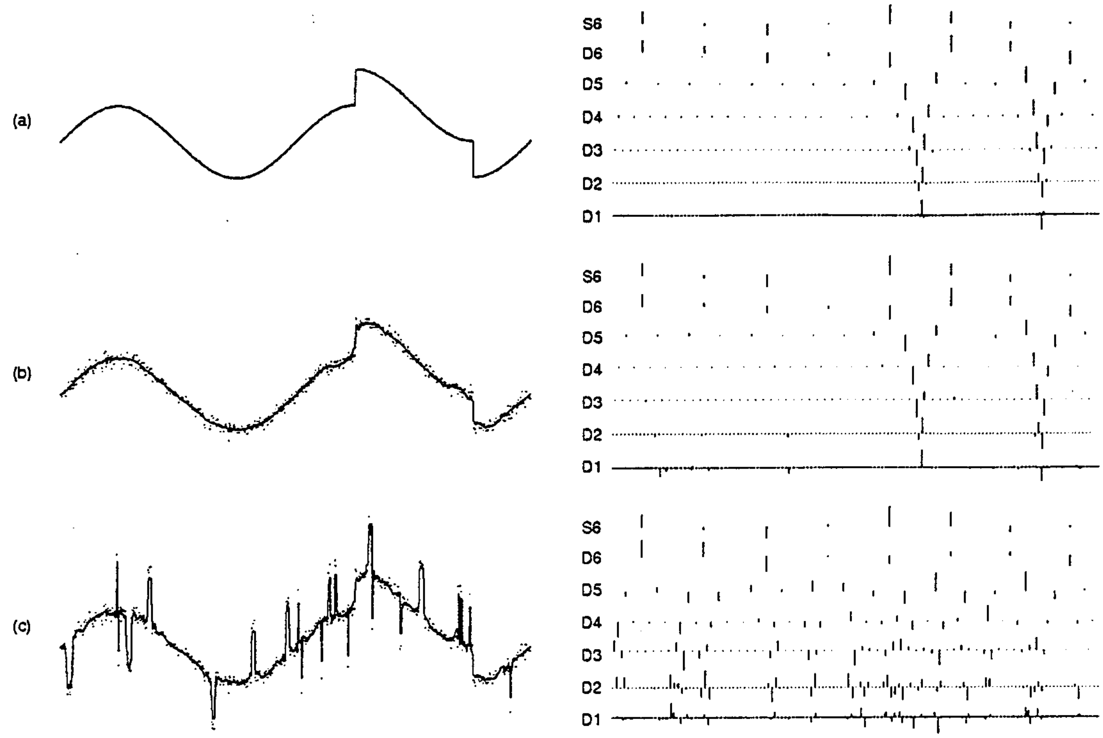


Figure 1: (a) The jumpsine signal (left plot) and its wavelet decomposition (right plot), (b) the signal contaminated by Gaussian noise (points), the WAVESHRINK smooth of the Gaussian contaminated data (solid line), and the wavelet decomposition corresponding to the WAVESHRINK smooth (c) The same as "(b)" except that the signal is also contaminated by impulsive patchy noise at random locations. While WAVESHRINK performs very well with Gaussian noise, WAVESHRINK is highly sensitive to outliers.

wavelet decomposition, shown in Figure 1(b). By shrinking the coefficients, we are able to remove most of the noise while still maintain the underlying signal, including the level shift. The data in figure 1(c) is obtained by further corrupting the signal with non-Gaussian impulsive outlier noise. The outliers are patches of fixed magnitude but random sign and patch length. The resulting WAVESHRINK smooth is very sensitive to the impulsive noise. The problem is that outliers are treated as local features by the WAVESHRINK procedure. Hence, like the level shift, outliers are preserved (see the corresponding wavelet decomposition).

One aim of our research is to broaden the scope of situations for which WAVESHRINK and related procedures are useful. To achieve this goal, we are developing a suite of algorithms for producing multiresolution and wavelet decompositions designed for signals which have noise distributions F_ϵ of the form

$$F_\epsilon = (1-\gamma)F + \gamma H \tag{4}$$

F is the "core" model, H is a "long tailed" outlier producing distribution, and γ is the fraction of contamination. We consider a variety of contamination models, emphasizing those which generate outliers occuring in patches [14].

The classical wavelet transform produces a sequence of approximations $\hat{f}_\ell(t)$ which are the projections of a signal $f(t)$ onto the basis formed by the collection of scaling functions $\phi_{\ell,k}(t) = 2^{-\ell/2}\phi(2^{-\ell}t - k)$. These projections minimize the L_2 norm

$$\| f(t) - \hat{f}_\ell(t) \|_2 \tag{5}$$

The L_2 norm, however, is well known to be very sensitive to outliers. In this section, we consider decompositions obtained by minimizing norms which are robust towards noise generated from models such as (4).

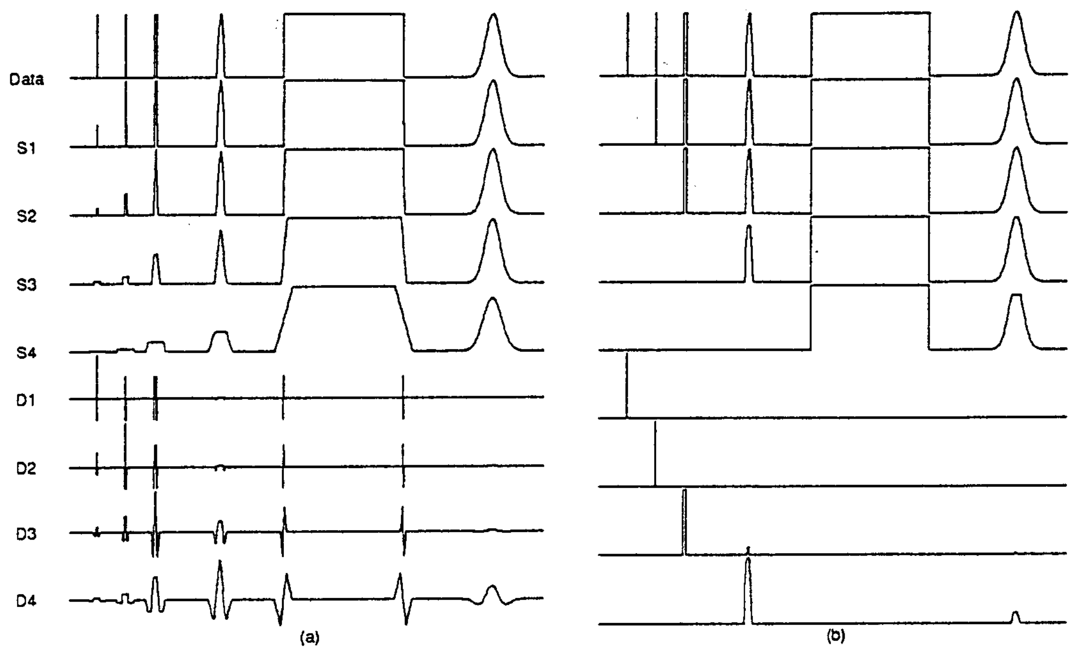


Figure 2: (a) L_2 multiresolution analysis using the Haar basis for an artificial signal consisting of outlier bursts, a smooth transient, a level shift, and smooth bump, and (b) L_1 multiresolution analysis for the same signal. The outlier bursts and transient are localized to the fine scales and concentrated in fewer coefficients for the L_1 decomposition.

3.1 L_1 Fitting and the Haar Basis

The L_1 norm is well known to be resistant towards outliers. As an example, we consider the analysis of an artificial signal using the Haar basis. For the Haar basis, the optimal L_2 and L_1 fits are given by the block mean and median respectively. Figure 2 compares the fits obtained by minimizing the L_2 norm (figure 2(a)) and L_1 norm (figure 2(b)). The original signal, given in the top plot, consists of three outlier bursts with different patch lengths, a smooth transient, a discontinuous level shift, and a smooth Gaussian kernel. The next four plots display the approximations $\mathrm{S}(\ell) \equiv \hat{f}_\ell(t)$ for $\ell = 1, 2, 3, 4$. The final four plots display the differences $\mathrm{D}(\ell) \equiv \hat{f}_{\ell-1}(t) - \hat{f}_\ell(t)$ for $\ell = 1, 2, 3, 4$. These differences are closely related to the "detail" coefficients in the classical wavelet transform. To ensure uniqueness of the L_1 fits, we use decimation by 3 in this examples.

This example illustrates two properties of L_2 and L_1 fits (in the decimation by 3 case):

P1: An outlier spike of length $\lfloor 3^\ell/2 \rfloor$ is isolated to levels $j = 0, 1, \ldots, \ell - 1$ for L_1 approximations $\hat{f}_\ell(t)$. By contrast, the outlier spikes are spread throughout the L_2 projections.

P2: The discontinuities and local transients are concentrated in fewer coefficients for the L_1 fits.

Hence, we can more easily remove outliers from the L_1 decomposition. Note also that the edges of the level shift are better preserved with the L_1 decomposition. The concentration of coefficients in the L_1 decomposition indicates that robust decompositions may have applications to data compression problems[4].

3.2 Smooth But Robust Fits

We can extend L_1 fitting to general wavelet bases and hence, obtain smoother projections. As an approximation, ℓ_1 fits can be used which are easily computed using using standard ℓ_1 regression techniques. However, L_1 fits are intrinsically non-smooth, and the so L_1 fits for general wavelet bases can still exhibit local roughness. We can obtain

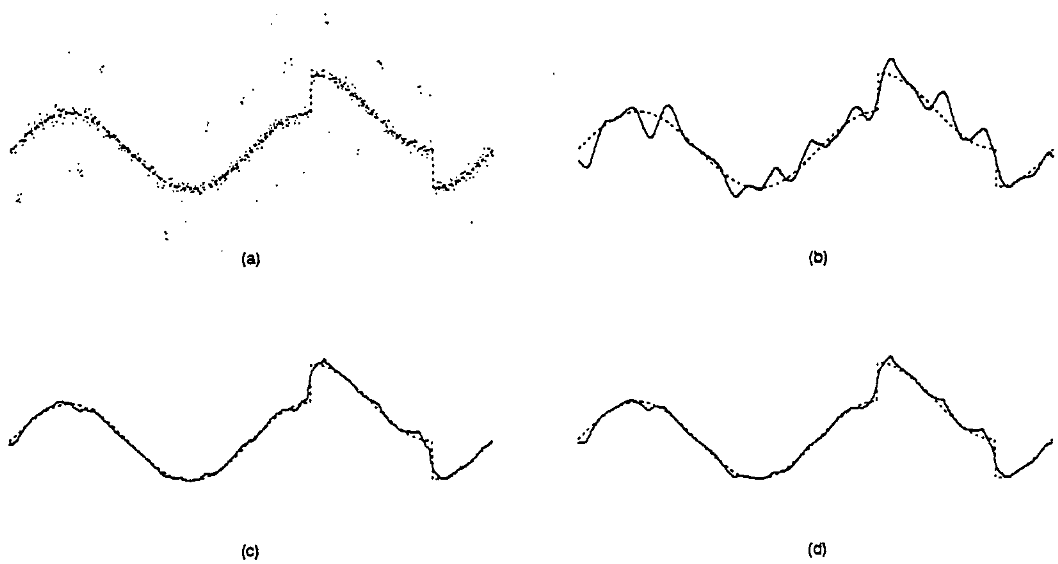


Figure 3: (a) The signal (line) contaminated by non-Gaussian noise (points), taken from in figure 1(c), (b) the L_2 projection at level 4 (decimation of the original signal by 2^4), (c) the L_1 fit at level 4, and (d) the hybrid fit at level 4 using the norm defined by (6). The hybrid analysis retains the smoothness of L_2 projection and the robustness of L_1 fit.

smooth but robust fits by using a hybrid loss function, such as

$$G(t) = \begin{cases} t^2 & \text{for } |t| \leq C \\ |t| & \text{for } |t| > C \end{cases} \tag{6}$$

Minimizing the norm defined by $G(t)$ retains the smoothness of L_2 projections and the robustness of L_1 fits.

To illustrate the difference between the different norms when using a smooth wavelet, we return to `jumpsine` example. Figure 3(a) plots the signal contaminated by non-Gaussian impulsive noise. Figure 3(b) gives the L_2 projection at level 4 (decimation of the original signal by 2^4) using the "least asymmetric" orthogonal wavelet with a filter of length 8 [3]. For illustrative purposes, a non-decimating shift invariant projection is performed [12]. Figures 3(c)-(d) give the corresponding fits based on minimizing the L_1 norm and the hybrid loss function respectively. The L_2 projection is significantly influenced by the outlier bursts. By contrast, L_1 and hybrid fits are relatively insensitive to the outliers. However, the hybrid fit is smoother and visually more appealing than the L_1 fit.

3.3 Computationally Unattractive

In general, the exact L_1 or hybrid approach is computationally too inefficient for practical use. Even in the Haar case, we do not retain the recursive filtering pyramid which makes the wavelet approach so attractive. It is our aim to mimic the robustness properties of this approach without sacrificing the computation efficiency of the discrete wavelet transform.

4 ROBUST SMOOTHER-CLEANER WAVELETS

The goal of robust smoother/cleaner wavelets is to produce a fast wavelet decomposition which is robust towards outliers. Smoother-cleaner wavelets behave like the classical L_2 wavelet transform for Gaussian signals, but prevent outliers and outlier patches from leaking into the wavelet coefficients at coarse levels (like L_1 wavelets). However, in contrast to the L_1 wavelets, algorithm is very fast with computational complexity $O(n)$.

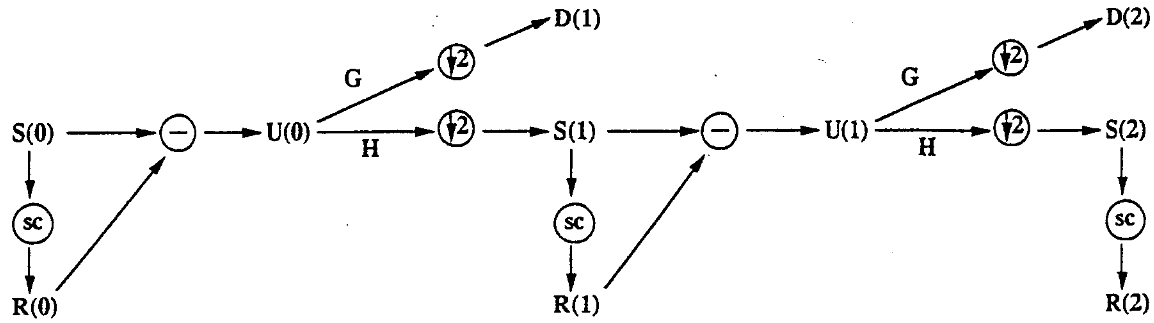


Figure 4: The robust smoother algorithm produces a pyramid decomposition with an extra component: the robust residual $R(\ell)$. For each multiresolution level, the low-pass coefficients $S(\ell)$ are first cleaned using a robust smoother cleaner, denoted by sc in the figure. The residuals are saved in the $R(\ell)$. The usual wavelet filters are then applied to the cleaned $S(\ell)$ to obtain $S(\ell+1)$ and $D(\ell+1)$.

4.1 Basic Algorithm

The basic idea of robust smoother/cleaner wavelets is simple: the smooth coefficients are preprocessed with a fast and robust smoother/cleaner. The procedure is illustrated in figure 4. As usual, we start with a set of wavelet coefficients $S(0)$. Then, for each multiresolution level, the signal is decomposed into three components:

1. A set of robust residuals $R(\ell-1)$, given by $R(\ell-1) = \delta_\lambda\left(S(\ell-1) - \hat{S}(\ell-1)\right)$ where δ_λ is the thresholder function (2) and $\hat{S}(\ell-1)$ is a robust smooth of $S(\ell-1)$ (e.g., running medians of 5). The threshold λ is chosen so that most of the robust residuals are zero.
2. The smooth wavelet coefficients $S(\ell)$ obtained by applying the usual low-pass/decimation wavelet filter H to the cleaned smooth coefficients $U(\ell-1) = S(\ell-1) - R(\ell-1)$.
3. The detail wavelet coefficients $D(\ell)$ obtained by applying the usual high-pass/decimation wavelet filter G to $U(\ell-1)$.

4.1.1 Choice of Wavelet Filters

The low-pass decomposition filters should be short in order to avoid leakage of outlier patches to the smooth coefficients. In general, the longer the low-pass filter, the more an outlier patch tends to get smeared when going from fine to coarse levels. The smearing is undesirable since it then is difficult to isolate and identify the outlier patch (as in the L_2 case). On the other hand, it is desirable to have longer filters to ensure sufficient smoothness with the underlying basis functions.

The "b-spline" biorthogonal wavelets[3] is one class of filters which satisfy both requirements: short filters can be used for decomposition while longer filters for reconstruction.

4.1.2 Choice of Robust Filter

The robust filter should be simple, computationally fast to compute, and have a very high breakdown point. Median filters are an attractive choice, and enjoy extensive usage in the engineering community. The width of the robust filter L should be as small as possible to provide minimal distortion of the underlying signal. However, L must be sufficiently big to prevent outlier patches from getting smeared in coarser scales.

In theory, for a low-pass wavelet filter of length M, smearing is prevented by using median filters of length $L \geq 2M+1$. In practice, we find that using median filters of length 5 or 7 is usually sufficient to avoid smearing for most types of wavelets.

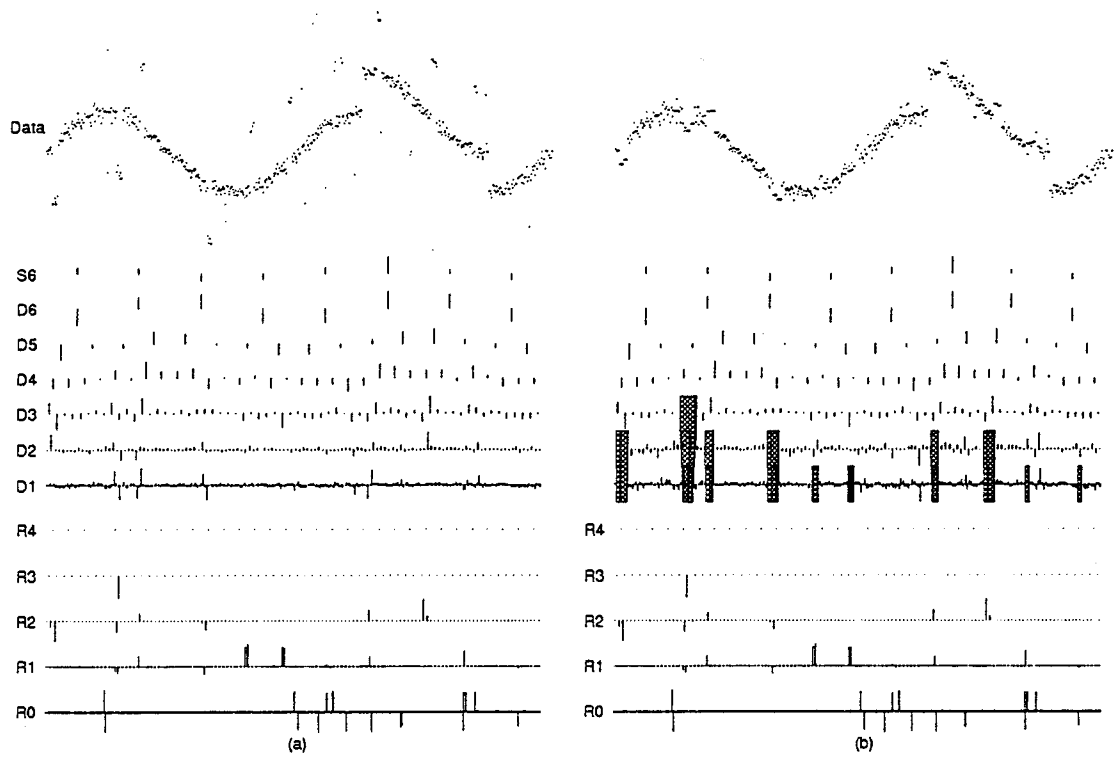


Figure 5: (a) Robust smoother-cleaner decomposition of the outlier contaminated `jumpsine` signal and (b) zero cones applied to the robust smoother decomposition in "(a)".

4.1.3 Setting the Robust Residual Threshold

The threshold λ determines the number of non-zero robust residuals. Setting λ too big will result in leakage of outliers into the signal and setting λ too small will cause distortion of the signal. We set λ so that an average of $100 \times p\%$ non-zero robust residuals remain after thresholding in the Gaussian case. The tuning parameter p is set to some small value (e.g., .01). A table for λ is obtained by simulation based on the Gaussian model. This value of λ is quite insensitive to the stochastic characteristics of the underlying signal.

4.2 Key Properties

- In the Gaussian noise case, the robust smoother-cleaner wavelet transform produces essentially the same decomposition as the classical wavelet transform. By design, only a small number of robust residuals are detected, and these will be small in magnitude (by virtue of the soft shrinkage function).
- If the data contains outliers and outlier patches, then the decomposition retains the dyadic equivalent of property **P1** for the L_1 decomposition of section 2: outliers patches of length $2^\ell + 2$ are isolated to wavelet coefficients at levels $j \leq \ell$.
- For certain outlier models of the type (4), it can be shown that WAVESHRINK, when applied appropriately to the robust smoother-clean wavelet decomposition (see below), can achieve a near optimality property similar to (3).

To illustrate the smoother-cleaner wavelet decomposition, we return to the `jumpsine` example of section 3. Figure 5(a) displays the smoother-cleaner wavelet decomposition for the outlier contaminated signal. In this example, we have used the Haar basis and median filter of length 5. The robust residuals correspond to the outlier bursts. Generally, the residuals R0 correspond to isolated outliers or pairs of outliers and the residuals R1-R3 correspond to longer outlier patches. Note that some longer patches are captured by pairs of residuals in R1 while other patches

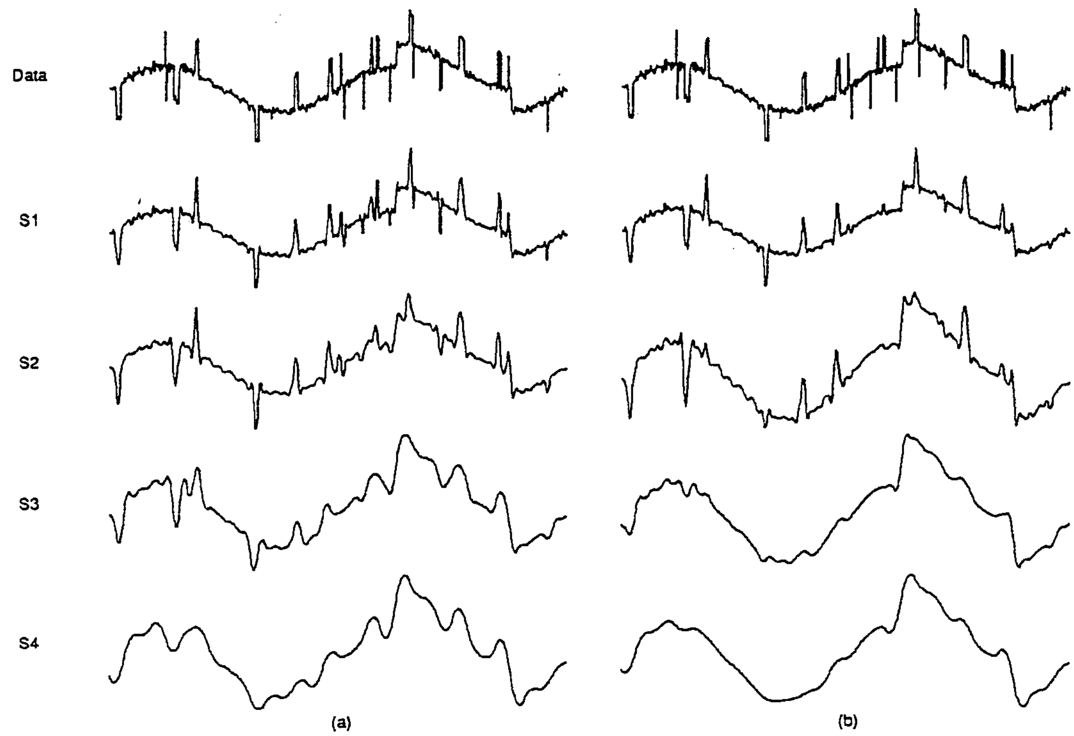


Figure 6: (a) Linear shrinkage applied to successively coarser levels for the discrete wavelet transform of the outlier corrupted jumpsine signal, and (b) linear shrinkage applied to robust smoother cleaner wavelet transform of the same data. The robust smoother-cleaner wavelet decomposition is superior to the classical wavelet transform for denoising by linear shrinkage.

are captured by single (large) residuals in R2 or even R3. The difference in the way in which outlier patches are represented in the robust residuals is due to decimation. Relative to the classical wavelet transform (see figure 1(c)), the robust smoother-cleaner wavelets have less leakage of the outlier patches into the coarse detail coefficients: compare the D3 and D4 coefficients.

The differences are even more evident when we look compare the robust and classical multiresolution analyses. Figure 6(a) gives a sequence of successively coarser estimates of the signal based on reconstructing from the $S1, S2, \ldots$ classical wavelet coefficients. This is equivalent to annihilating all detail coefficients at scales D1, (D1, D2), (D1, D2, D3), ... Figure 6(b) gives the analogous sequence based on the robust smoother-cleaner wavelet coefficients. The robust reconstructions are much less sensitive to the outlier bursts. The $S4$ robust fit closely mimics the L_1 and Huber fits of figure 3.

4.3 Combining with Waveshrink

Non-linear shrinkage, such as that used by the WAVESHRINK procedure, can outperform linear shrinkage when the signal contains local features which we want to preserve in the finest detail wavelet coefficients. For example, note how the jump in the signal becomes blurred at the coarser scales in figure 6. By contrast, the WAVESHRINK estimate for the same signal contaminated only by Gaussian data preserves the jump: see figure 1(b). In this section, we discuss how the WAVESHRINK algorithm can be applied to data with outliers by using the robust smoother-cleaner basis.

The simplest procedure is to discard the robust residuals and to use WAVESHRINK on the wavelet coefficients. If the data contains only isolated or pairs of outliers, this procedure will generally work. However, if the data contains longer bursts of outliers, then this procedure is likely to breakdown. While the robust smoother-cleaner prevents outliers from leaking into the coarse scale detail and smooth coefficients, it does not prevent outliers from patches

leaking into fine scale detail coefficients. In figure 5(a), we see that the largest D1 and D2 coefficients are associated with outlier patches. Hence, applying WAVESHRINK to the decomposition of figure 5(a) will result in some leakage of the noise into the signal.

To get around this problem, we provide two solutions: selective annihilation of coefficients using "zero cones" and a "clean and repeat" procedure. These are discussed below.

4.3.1 Zero Cones

The basic principles behind zero cones are that

- Every patch of outliers will eventually be detected by some robust residual.
- An influence cone can be constructed indicating which detail coefficients at levels ℓ, $\ell-1, \ldots, 1$ are computed from the outlier patch associated with a robust residual at level ℓ

By shrinking all coefficients in the influence cone to zero (or to a suitable threshold), we can ensure that we bound the influence of any large detail coefficients associated with an outlier patch. We denote this as the "zero cone" procedure, since all coefficients are annihilated in a specified cone. In practice, to avoid artifacts caused by over-shrinking, we use zero cones for only moderate or large robust residuals.

Figure 5(b) displays the result of applying the zero cone procedure to the smoother-cleaner wavelet decomposition. The data at top is the result of reconstructing from the zero cone wavelet decomposition (without the robust residuals). The influence of the outlier patches has been almost entirely removed. The zero cones are superimposed on the plot of the decomposition. The largest detail coefficients at levels D1 and D2 have been set to zero by the zero cones: compare with figure 5(a). The remaining large coefficients at these levels correspond to the level shift.

The smoother-cleaner wavelets combined with the zero-cone procedure achieves our goal: large detail coefficients associated with outlier patches are annihilated but those associated with features are preserved. The result of applying the WAVESHRINK procedure to the zero-cone wavelets in this example is given in figure 7(a). The estimated signal (solid blocky line) faithfully tracks the true signal (dashed line), preserving the discontinuity. The outlier patches result in minimal leakage into the signal. The estimated signal is blocky since the Haar basis is used.

We remark that the zero cone procedure is especially attractive when combined with the power of an modern graphics workstation. Using the mouse, the user can interactively select cones which correspond to suspected outlier patches. In this context, zero cones can be applied both to the robust smoother-cleaner transform as well as the classical wavelet transform. See section 5 for further discussion of software.

4.3.2 Clean and Repeat

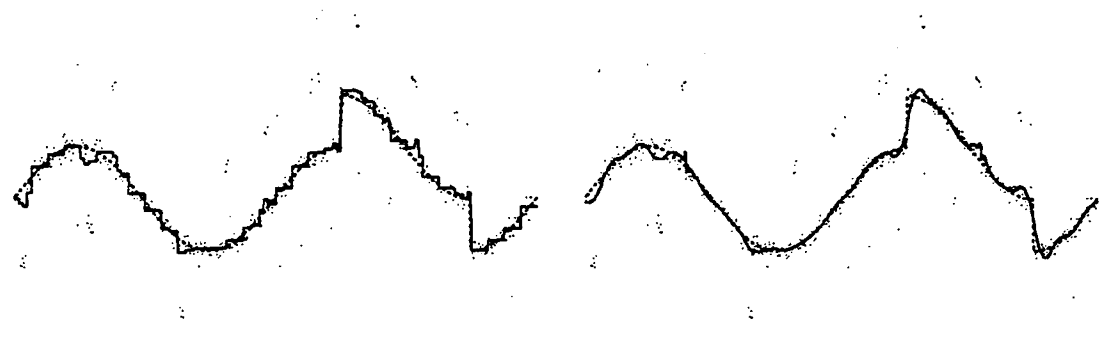
While zero cones are intuitive and computationally very fast, they have the drawback that the cones can get very wide for general wavelet bases. In addition, zero cones require careful tracking of the indices taking into account the various possible configurations which result from decimation. A very simple alternative to the zero cones procedure which is almost as fast and empirically produces as good or better results is the "clean and repeat" procedure:

[0] Initialize with $j = 0$ and $\hat{X}_0(t) = X(t)$.

[1] Apply the robust smoother-cleaner wavelet decomposition to a data sequence $\hat{X}_j(t)$. If the number and magnitude of the robust residuals is sufficiently small, then quit.

[2] Reconstruct without the robust residuals to obtain a clean data sequence $\hat{X}_{j+1}(t)$.
Set $j = j + 1$ and go to step 1.

The basic idea is that the repeated applications of the the robust smoother-cleaner will capture any outliers which leak into the detail coefficients in previous applications. In practice, only *two passes* of the robust smoother-cleaner operations are necessary to clean the data. The resulting decomposition contains the usual wavelet coefficients plus $j-1$ sets of robust residuals.



(a) (b)

Figure 7: (a) The estimated signal obtained by the WaveShrink algorithm applied to the smoother-cleaner wavelet decomposition combined with the zero cone procedure and (b) the estimated signal using WaveShrinkapplied to decomposition obtained from the clean and repeat procedure.

Figure 7(b) shows the result of applying WaveShrink to the "clean and repeat" decomposition of the outlier contaminated `jumpsine` data. The "b-spline" biorthogonal wavelet ${\psi_{1,5}}^3$ is used for this example. The smooth is very similar to the estimate obtained by the zero cone procedure. The main difference is that we have used a smooth basis function instead of the Haar basis.

4.4 Application: Denoising Radar Glint Noise Data

We now apply the robust denoising procedures to radar glint noise data. The original noisy signal, which is the angle of the target in degrees, is displayed in Figure 8(a) The true signal is a low-frequency oscillation about 0°. The signal contains a number of glint spikes, causing the apparent signal to be different from the true signal by as much as 150°.

Figure 8(b) compares denoising with linear shrinkage of wavelets (dashed line) to denoising with WaveShrink combined with robust smoother-cleaner wavelets obtained by the clean and repeat procedure (solid line). The linear shrinkage is based on annihilating all detail coefficients of the classical wavelet transform at levels $\ell = 1, 2, 3, 4$. While linear shrinkage estimate is smooth, it is still somewhat sensitive to the glint spikes. By contrast, the clean and repeat procedure is quite resistant to the glint spikes but effectively tracks the sudden changes in the series.

5 DISCUSSION

Our research was motivated by a problem central in time series analysis: how to extract non-stationary signals which may have abrupt changes, such as level shifts, in the presence of impulsive outlier noise. A variety of techniques have been employed to deal with the problem, such as robust Kalman filtering[10, 15] and iterative outlier identification[2]. Our research indicates that a wavelet approach is an attractive alternative, offering a very fast algorithm with good theoretical properties.

Wavelets are not an appropriate basis for analysis of all types of signals. Researchers have offered various alternative bases, such as wavelet packets and local cosine bases. In this paper, we have presented some variations of wavelet analysis for data which contains impulsive outlier noise. See below for a discussion of other related research efforts.

A rich software environment is needed to support the rapid proliferation of wavelet-like techniques for analyzing data require. In a complimentary research project, we are developing S+Wavelets, an object-oriented toolkit for

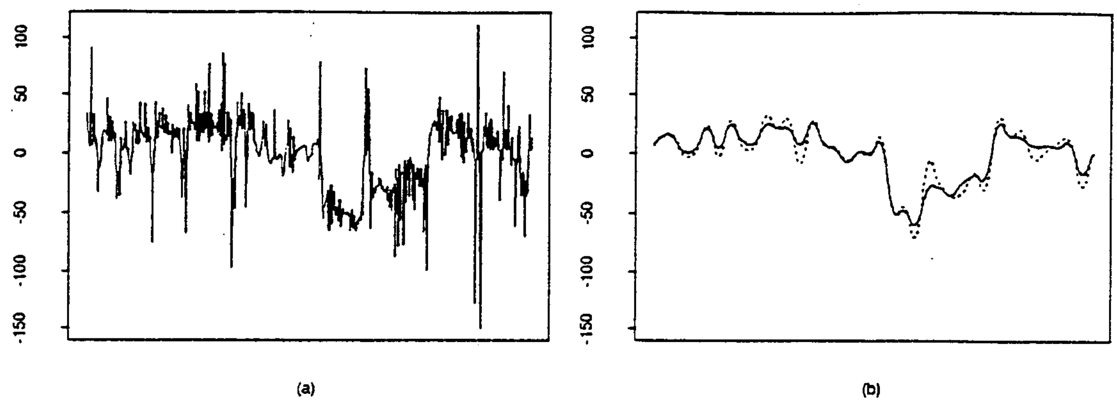


Figure 8: (a) Radar glint noise data in degrees, and (b) denoising by linear shrinkage of wavelets (dashed line) compared with denoising by WAVESHRINK combined with robust smoother-cleaner wavelets obtained by the clean and repeat procedure (solid line).

wavelet analysis. The robust algorithms discussed in this paper are embedded in this toolkit. S+WAVELETS is briefly discussed below.

5.1 Related Research

Robust multiresolution decompositions based on median filtering have been developed elsewhere[4, 11] and applied to problems such as analysis of mammograms[16]. We are developing other new algorithms for robust wavelet analysis. In one approach, we develop a nonlinear triadic refinement scheme in which the wavelet coefficients are possibly nonlinear combinations of finitely many block medians at a given scale. These wavelets are nonlinear cousins of the Deslaurier-Dubuc[9] interpolation scheme. In another approach, we combine approximate conditional mean smoothers[13] with wavelets implemented using IIR filters.

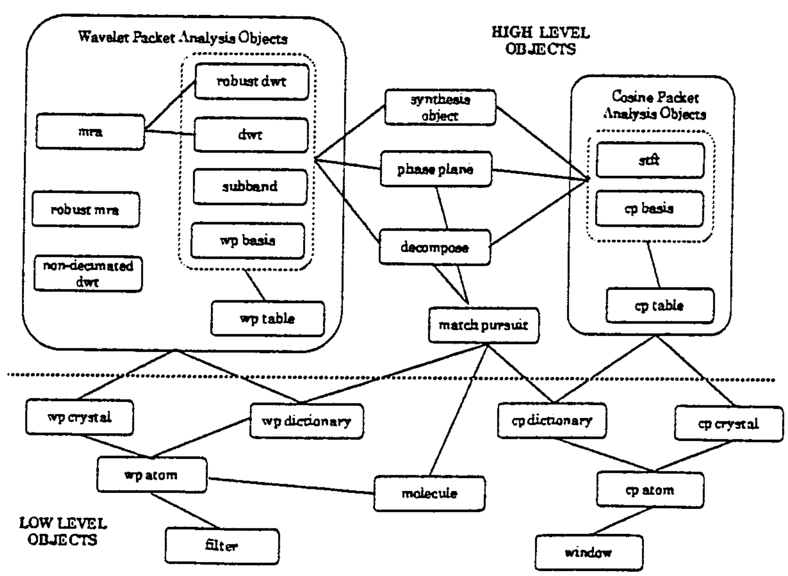


Figure 9: Overview of S+WAVELETS. Classes of objects are represented by the nodes. The arcs represent conceptual links between objects.

5.2 S+Wavelets Object-Oriented Toolkit

S+WAVELETS is an extensible ***object-oriented language*** for wavelet analysis. S+WAVELETS is based on the S language for data analysis, graphics, and statistics [1]. An overview of S+WAVELETS is given in figure 9. Classes of objects are represented by the nodes. The arcs represent conceptual links between objects. The toolkit offers an array of basic building blocks, including waveform "atoms" and "crystals", wavelet filters, and time frequency dictionaries. These low-level objects are used to construct the higher-level objects, such as a wavelet transform or multiresolution analysis. The high-level objects are organized into a class hierarchy, utilizing inheritance for sharing behavior. In addition to the new robust algorithms, S+WAVELETS includes the discrete wavelet transform (one-dimensional and two-dimensional), wavelet packets and local cosine bases, non-decimating wavelets, a variety of graphical displays, and careful treatment of boundary related issues.

6 ACKNOWLEDGEMENTS

This research was supported by the Office of Naval Research.

7 REFERENCES

[1] R. A. Becker, J. C. Chambers, and A. R. Wilks. *The New S Language: An Programming Environment for Data Analysis and Graphics*. Wadsworth, 1988.

[2] I. Chang, G. C. Tiao, and C. Chen. Estimation of time series parameters in the presence of outliers. *Technometrics*, 30:193–204, 1988.

[3] I. Daubechies. *Ten lectures on wavelets*. Society for industrial and applied mathematics, Philadelphia, PA, 1992.

[4] Ronald A. DeVore, Bjorn Jawerth, and Bradly J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory*, 38(2):719–746, 1992.

[5] David L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In Ingrid Daubechies, editor, *Proceedings of the Symposia in Applied Mathematics*. American Mathematical Society, 1993.

[6] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[7] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Technical report, Stanford University, 1992.

[8] David L. Donoho and Iain M. Johnstone. Minimax estimation via wavelet shrinkage. Technical Report 402, Stanford University, 1992.

[9] S. Dubuc. Interpolation through an iterative scheme. *J. Math. Anal. and Appl.*, 114:185–204, 1986.

[10] P. J. Harrison and C. F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society, Series B*, 38:205–247, 1976.

[11] H. Longbotham. A class of robust nonlinear filters for signal decomposition and filtering utilizing the Haar basis. In *ICASSP-92*, volume 4. IEEE Signal Processing Society, March 1992.

[12] Stéphane Mallat and Sifen Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710–732, 1992.

[13] R. D. Martin. Approximate conditional-mean type smoothers and interpolators. In T. Basser and M. Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*, pages 147–176. Academic Press, New York, 1979.

[14] R. D. Martin and V. J. Yohai. Influence curves for time series. *The Annals of Statistics*, 11:781–818, 1986.

[15] C. J. Masreliez and R. D. Martin. Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic Control*, AC-22:361–371, 1977.

[16] Walter B. Richardson. Nonlinear filtering and multiscale texture discrimination for mammograms. Technical report, University of Texas, San Antonio, TX 78249, 1993.